

Statistical Modelling of Counts with a Simple Integer-Valued Bilinear Process



Isabel Pereira and Nélia Silva

Abstract The aim of this work is the statistical modelling of counts assuming low values and exhibiting sudden and large bursts that occur randomly in time. It is well known that bilinear processes capture these kind of phenomena. In this work the integer-valued bilinear INBL(1,0,1,1) model is discussed and some properties are reviewed. Classical and Bayesian methodologies are considered and compared through simulation studies, namely to obtain estimates of model parameters and to calculate point and interval predictions. Finally, an empirical application to real epidemiological count data is also presented to attest for its practical applicability in data analysis.

1 Introduction

In the analysis of stationary integer-valued time series the class of INARMA models plays a central role. However, such models are unlikely to provide a sufficiently broad class capable of accurately capturing features often exhibited by data sets such as sudden burst of large values. For that purpose and using the concept of thinning operator, introduced by [12], conventional bilinear models can be adapted to the integer case leading to the class of integer-valued bilinear models. Doukhan et al. [3] proposed the first-order INBL(1,0,1,1) model

$$X_t = \alpha \circ X_{t-1} + \beta \circ (\epsilon_{t-1} X_{t-1}) + \epsilon_t. \quad (1)$$

I. Pereira (✉) · N. Silva

Departamento de Matemática and CIDMA, University of Aveiro, Aveiro, Portugal
e-mail: isabel.pereira@ua.pt; neliasilva@ua.pt

where the thinning operators “ $\alpha \circ$ ” and “ $\beta \circ$ ”¹ are mutually independent, $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of i.i.d. non-negative integer-valued random variables with finite mean and finite variance, independent of the operators. Doukhan et al. [3] derived conditions guaranteeing strictly and second-order stationarities of INBL(1,0,1,1) model. Drost et al. [4] also provided sufficient conditions for the existence of higher order moments of X_t , considering the superdiagonal INBL(p, q, m, n). One step towards the application of bilinear models to real data sets is the estimation of parameters. Considering Poisson thinning operators [3] have obtained moments estimators and derived their asymptotic distribution. In contrast, Bayesian analysis of INBL has not received much attention in the literature neither diagnostic analysis.

In this paper we consider the INBL(1,0,1,1) given in (1), with the following assumptions: the operators “ $\alpha \circ$ ” and “ $\beta \circ$ ” are mutually independent such as $\alpha \circ X_{t-1} | X_{t-1} \sim Bi(X_{t-1}, \alpha)$, $\beta \circ (\epsilon_{t-1} X_{t-1}) | X_{t-1}, \epsilon_{t-1} \sim Bi(\epsilon_{t-1} X_{t-1}, \beta)$ and $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of i.i.d. Poisson random variables with mean λ , independent of the operators.

The class of stationary models defined in (1) is useful for representing time series that assume low values with high probability and exhibit sudden bursts of large values that occur randomly in time, hence can produce heavy-tailed data. As an illustration of this kind of data we present in Fig. 1 two time series of count data in epidemiology, originally studied by [3]. Data consist of the weekly number of E. coli infections and meningitis cases, both starting in January 1990 and corresponding to 143 observations for each series. Counts are typically small, skewed and both series contain a large quantity of zeros. Hereafter the weekly number of E.coli infections and weekly number of meningitis cases are denoted by the E.coli data and meningitis data, respectively.

In time series analysis we usually are interested in estimating the underlying model and in predictive capabilities of that model. Thus, the aim of this study is to establish a comparison between classical and Bayesian approaches in order to

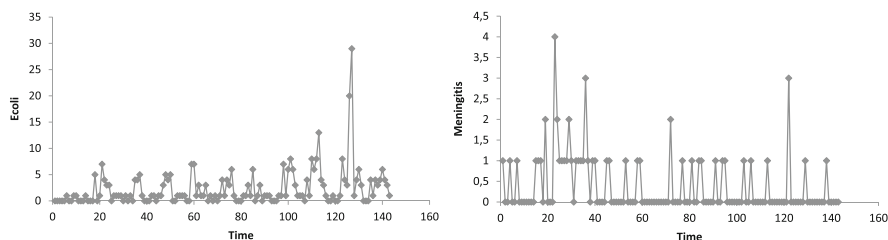


Fig. 1 Time series plots for E. coli data (*left*) and meningitis data (*right*)

¹Steutel and van Harn operator “ $\phi \circ$ ” is defined by $\phi \circ X = \sum_{i=1}^X Y_i$ where $\{Y_i\}$, $i = 1, \dots, X$, is a sequence of independent and identically distributed (i.i.d.) counting random variables with mean ϕ and X is a non-negative integer-valued random variable, independent of Y . If Y_i is a Bernoulli random variable, we have the binomial thinning operator.

conduct inference for model parameters and obtain predictions for future values. The rest of the paper is organized as follows. Classical and Bayesian methodologies are presented to obtain estimates of the model parameters in Sect. 2 and forecasting is addressed in Sect. 3. The performance of the above procedures is illustrated through a simulation study in Sect. 4. Section 5 provides applications of this model to real data sets. Finally, Sect. 6 contains some concluding remarks.

2 Parameters Estimation

One step forward the application of INBL models in practice is the estimation of their corresponding parameters $\theta = (\theta_1, \theta_2, \theta_3) = (\alpha, \beta, \lambda)$. The classical estimators studied are grouped according to two broad categories: regression-based and likelihood-based estimators. Furthermore, Bayesian estimation is also considered. In any estimation procedure, it is required to estimate the r.v.s ϵ_t since they are not observable. From (1), the innovations ϵ_t can be recursively calculated through $\epsilon_t = X_t - \alpha \circ X_{t-1} + \beta \circ (\epsilon_{t-1} X_{t-1})$, $t = 1, \dots, n$, using an initial value for ϵ_1 .

2.1 Conditional Least Squares Estimators

The CLS-estimators of θ are obtained by minimizing

$$Q(\theta) = \sum_{t=2}^n [X_t - E(X_t | X_{t-1}, \epsilon_{t-1})]^2 = \sum_{t=2}^n [X_t - \alpha X_{t-1} - \beta X_{t-1} \epsilon_{t-1} - \lambda]^2,$$

yielding to the following expressions for the parameters estimators

$$\begin{aligned} \hat{\beta}_{\text{CLS}} &= \frac{S_{t-1} S_{t-1;(t-1, \epsilon-1)} - S_{t-1;t-1} S_{t-1;(t-1, \epsilon-1)}}{S_{t-1;(t-1, \epsilon-1)}^2 - S_{t-1, \epsilon-1} S_{t-1;t-1}}, \\ \hat{\alpha}_{\text{CLS}} &= \frac{(n-1) S_{t-1;t-1} - \hat{\beta}_{\text{CLS}} S_{t-1;(t-1, \epsilon-1)}}{S_{t-1;t-1}}, \\ \text{and} \\ \hat{\lambda}_{\text{CLS}} &= \frac{\sum_{t=2}^n X_t - \hat{\alpha}_{\text{CLS}} \sum_{t=2}^n X_{t-1} - \hat{\beta}_{\text{CLS}} \sum_{t=2}^n X_{t-1} \epsilon_{t-1}}{n-1}, \end{aligned}$$

with $i, j = 0, 1$,

$$\begin{aligned} \bar{X}_j &= \frac{1}{n-1} \sum_{t=2}^n X_{t-j} \\ S_{t-i, t-j} &= \sum_{t=2}^n (X_{t-i} - \bar{X}_i)(X_{t-j} - \bar{X}_j) \\ S_{t-i; (t-1, \epsilon-1)} &= \sum_{t=2}^n (X_{t-i} - \bar{X}_i)(X_{t-1} \epsilon_{t-1} - \bar{X}_1 \bar{\epsilon}_1), \quad \bar{\epsilon}_1 = \frac{1}{n-1} \sum_{t=2}^n \epsilon_{t-1}, \\ S_{t-1, \epsilon-1} &= \sum_{t=2}^n (X_{t-1} \epsilon_{t-1} - \bar{X}_1 \bar{\epsilon}_1)^2. \end{aligned}$$

2.2 Conditional Maximum Likelihood Estimators

For fixed values of x_1 and ϵ_1 , the conditional log-likelihood function for the INBL(1,0,1,1) model is given by

$$l(\boldsymbol{\theta}) := \ln L(\mathbf{x}_n; \boldsymbol{\theta} | x_1, \epsilon_1) = \sum_{t=2}^n \ln(p(x_t | x_{t-1}, \epsilon_{t-1})),$$

with $\mathbf{x}_n = (x_1, \dots, x_n)$ and transition probabilities

$$\begin{aligned} p(x_t | x_{t-1}, \epsilon_{t-1}) &= \sum_{k=0}^{m_t} \exp(-\lambda) \frac{\lambda^{x_t-k}}{(x_t-k)!} \\ &\times \sum_{j=M_t}^{m_{tk}} \binom{x_{t-1}}{j} \alpha^j (1-\alpha)^{x_{t-1}-j} \binom{x_{t-1}\epsilon_{t-1}}{k-j} \beta^{k-j} (1-\beta)^{x_{t-1}\epsilon_{t-1}-k+j}, \end{aligned}$$

$m_t = \min(x_t, x_{t-1} + x_{t-1}\epsilon_{t-1})$, $M_t = \max(0, k - x_{t-1}\epsilon_{t-1})$ and $m_{tk} = \min(k, x_{t-1})$, since the r.v. $X_t | x_{t-1}, \epsilon_{t-1}$ is the convolution between binomial distributions with parameters (X_{t-1}, α) and $(\epsilon_{t-1} X_{t-1}, \beta)$, respectively, and Poisson distribution with parameter λ . The CML-estimators are obtained by maximizing the conditional log-likelihood function. Due to the complexity of log-likelihood expression it is not possible to give explicit forms to the CML-estimators of α , β , and λ , thus it is necessary to use numerical procedures. The initial estimates required by such numerical procedures can be obtained by the method of least squares or using moment estimates given by [3] procedure.

2.3 Bayesian Approach

To implement the Bayesian version of the INBL(1,0,1,1) model we need to consider prior distributions for the parameters. Thus, for the parameters $0 < \alpha, \beta < 1$ we choose Beta priors with hyperparameters (a, b) and (c, d) , respectively while for the positive parameter λ we choose a Gamma prior with hyperparameters (e, f) . These priors are traditionally used for the PoINAR(1) by [11].

Given the particular sample \mathbf{x}_n , the updated information about $\boldsymbol{\theta}$ is expressed through Bayes theorem by the posterior distribution $\pi(\boldsymbol{\theta} | \mathbf{x}_n)$ given by

$$\pi(\boldsymbol{\theta} | \mathbf{x}_n) = \frac{L(\mathbf{x}_n; \boldsymbol{\theta} | x_1, \epsilon_1) \pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} L(\mathbf{x}_n; \boldsymbol{\theta} | x_1, \epsilon_1) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto L(\mathbf{x}_n; \boldsymbol{\theta} | x_1, \epsilon_1) \pi(\boldsymbol{\theta}),$$

with $\pi(\boldsymbol{\theta})$ representing the prior distribution.

Assuming independence assumptions on the parameters the posterior distribution is given by

$$\pi(\boldsymbol{\theta}|\mathbf{x}_n) \propto \lambda^{e-1} e^{-f\lambda} \alpha^{a-1} (1-\alpha)^{b-1} \beta^{c-1} (1-\beta)^{d-1} \left(\prod_{t=2}^n \sum_{k=0}^{m_t} \exp(-\lambda) \frac{\lambda^{x_t-k}}{(x_t-k)!} \sum_{j=M_t}^{m_{tk}} \binom{x_{t-1}}{j} \alpha^j (1-\alpha)^{x_{t-1}-j} \binom{x_{t-1}\epsilon_{t-1}}{k-j} \beta^j (1-\beta)^{x_{t-1}\epsilon_{t-1}-k+j} \right),$$

with $\lambda > 0$, $0 < \alpha, \beta < 1$, $m_t = \min(x_t, x_{t-1} + x_{t-1}\epsilon_{t-1})$, $M_t = \max(0, k - x_{t-1}\epsilon_{t-1})$, and $m_{tk} = \min(k, x_{t-1})$.

Thus given the complexity of the posterior distribution, Markov Chain Monte Carlo (MCMC) techniques are required for sampling purposes. For the simulations we need the full conditional distributions for each parameter θ_i , denoted by $\pi(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{x}_n)$, which is the posterior distribution of θ_i conditional on all other parameters and the data \mathbf{x}_n . The full conditional distributions of α , β , and λ are, respectively:

$$\pi(\alpha|\beta, \lambda, \mathbf{x}_n) \propto \alpha^{a-1} (1-\alpha)^{b-1} L(\mathbf{x}_n; \boldsymbol{\theta}|x_1, \epsilon_1),$$

$$\pi(\beta|\alpha, \lambda, \mathbf{x}_n) \propto \beta^{c-1} (1-\beta)^{d-1} L(\mathbf{x}_n; \boldsymbol{\theta}|x_1, \epsilon_1)$$

and

$$\pi(\lambda|\alpha, \beta, \mathbf{x}_n) \propto \lambda^{e-1} e^{-f\lambda} L(\mathbf{x}_n; \boldsymbol{\theta}|x_1, \epsilon_1).$$

From the above expressions it is easy to conclude that the full conditional distributions will not be standard distributions and therefore a componentwise Metropolis- Hastings algorithm is used, particularly the Adaptive Rejection Metropolis Sampling (ARMS), as described in [7]. After having generated samples $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ sample central tendency measures are used to estimate the model parameters.

3 Prediction Future Observations

In this section we consider the problem of predicting the values of X_{n+h} , $h \in N$ for INBL (1,0,1,1) process based on the observed series up to time n . The usual way of producing forecasts is via the conditional predictive distribution and the most common procedure for obtaining predictions in time series models is to use conditional expectations, since we pretend to minimize the mean square error. Throughout this section we consider $\mathbf{B}_n = \{X_1, \dots, X_n; \epsilon_1, \dots, \epsilon_n\}$.

3.1 Classical Approach

3.1.1 Point Predictions

For $h \geq 1$ the h -step-ahead predictor can be obtained in a recursive way through

$$\hat{X}_{n+h} = E(X_{n+h}|\mathbf{B}_n) = \alpha \hat{X}_{n+h-1} + \beta E(\epsilon_{n+h-1} X_{n+h-1} | \mathbf{B}_n) + \lambda \quad (2)$$

with

$$\begin{aligned} E(\epsilon_{n+h-1} X_{n+h-1}) &= E[\epsilon_{n+h-1}(\alpha \circ X_{n+h-2} + \beta \circ (\epsilon_{n+h-2} X_{n+h-2}) + \epsilon_{n+h-1}) | \mathbf{B}_n] \\ &= \alpha \lambda \hat{X}_{n+h-2} + \lambda \beta E(\epsilon_{n+h-2} X_{n+h-2}) + \lambda^2 + \lambda. \end{aligned}$$

For the particular cases $h = 1$ and $h = 2$ we have, respectively,

$$\begin{aligned} \hat{X}_{n+1} &= \alpha X_n + \beta(X_n \epsilon_n) + \lambda, \\ \hat{X}_{n+2} &= \alpha^2 X_n + \alpha \beta(X_n \epsilon_n) + \alpha \lambda + \alpha \beta \lambda X_n + \beta^2 \lambda(X_n \epsilon_n) + \beta(\lambda^2 + \lambda) + \lambda. \end{aligned}$$

We can easily prove that

$$\lim_{h \rightarrow +\infty} \hat{X}_{n+h} = \frac{\beta \lambda^2 + \beta \lambda + \lambda(1 - \lambda \beta)}{(1 - \lambda \beta)(1 - \alpha) - \alpha \beta \lambda}. \quad (3)$$

Since these predictors based on conditional expectation hardly produce integer-valued forecasts, we can alternatively use the median of h -step-ahead conditional distribution of $X_{n+h} | \mathbf{B}_n$, denoted by \hat{M}_{n+h} , to obtain coherent predictions of X_{n+h} , as suggested in [5].

3.1.2 Prediction Intervals for One-Step-Ahead Observation

The one-step-ahead prediction error

$$e_{n+1} = X_{n+1} - \hat{X}_{n+1} = X_{n+1} - \hat{\alpha} X_n - \hat{\beta}(X_n \epsilon_n) - \hat{\lambda}$$

is a discrete variable with probability function

$$\begin{aligned} P(e_{n+1} = x_{n+1} - g) &= P(X_{n+1} = x_{n+1} | \mathbf{B}_n) = f(x_{n+1} | \mathbf{B}_n; \theta) \\ &= \sum_{k=0}^{m_1} \exp(-\lambda) \frac{\lambda^{x-k}}{(x-k)!} \times \sum_{l=M_1}^{m_2} \binom{x_n}{l} \alpha^l (1-\alpha)^{x_n-l} \times \binom{x_n \epsilon_n}{k-l} \beta^{k-l} (1-\beta)^{x_n \epsilon_n - k + l} \end{aligned}$$

where $g = \hat{\alpha} X_n + \hat{\beta}(X_n \epsilon_n) + \hat{\lambda}$. Hence the γ level confidence interval for X_{n+1} is given by: $(\hat{X}_{n+1} + l_1, \hat{X}_{n+1} + l_2)$ where l_1 is the largest value of e_{n+1} such as $P(e_{n+1} \leq l_1) \leq (1 - \gamma)/2$ and l_2 the smallest value of e_{n+1} such as $P(e_{n+1} \leq l_2) \geq (1 + \gamma)/2$.

3.2 Bayesian Predictions

The Bayesian predictive probability function is based on the assumption that both the future observation X_{n+h} and θ are unknown. The conditional distribution of X_{n+h} given \mathbf{B}_n which can be viewed as containing all the accumulated information about the future, represents the h -step-ahead Bayesian posterior predictive distribution. It is defined by

$$\pi(x_{n+h}|\mathbf{B}_n) = \int_{\Theta} f(x_{n+h}|\mathbf{B}_n; \theta) \pi(\theta|\mathbf{x}_n) d\theta,$$

with $\theta \in \Theta$ being the vector of unknown parameters, $\pi(\theta|\mathbf{x}_n)$ the posterior density of θ , and $f(x_{n+h}|\mathbf{B}_n; \theta)$ the (classical) predictive distribution.

3.2.1 Point Predictions for the Future Observation

In the particular case of $h = 1$ the one-step-ahead Bayesian predictive distribution is given by

$$\pi(x_{n+1}|\mathbf{B}_n) = \int_{\alpha} \int_{\beta} \int_{\lambda} f(x_{n+1}|\mathbf{B}_n; \theta) \pi(\theta|\mathbf{x}_n) d\alpha d\beta d\lambda,$$

with $m_1 = \min(x_{n+1}, x_n + \epsilon_n x_n)$, $m_2 = \min(x_n, k)$, $M_1 = \max(0, k - x_n \epsilon_n)$.

The Bayesian predictor of X_{n+1} can be obtained by any location measure of the predictive distribution. Its complexity does not allow work with it directly. However we can adapt to the integer case the Tanner composition method (as reported in [13]), to get an estimate of X_{n+h} using the sample mean or sample median of the generated values $(X_{n+h,1}, \dots, X_{n+h,m})$. Similarly to the classical case we can use the recursive expression

$$\begin{aligned} E(X_{n+h}|\mathbf{B}_n) &= E[E(X_{n+h}|\theta, \mathbf{B}_n)] = E\left[E\left(\alpha \circ X_n + \beta \circ (\epsilon_n X_n) + \epsilon_{n+1} \mid \theta, \mathbf{B}_n\right)\right] \\ &= \hat{\alpha}_B \hat{X}_{n+h-1} + \hat{\beta}_B E(\epsilon_{n+h-1} X_{n+h-1} | \mathbf{B}_n) + \hat{\lambda}_B \end{aligned} \quad (4)$$

where $\hat{\alpha}_B$, $\hat{\beta}_B$, and $\hat{\lambda}_B$ are the Bayesian estimates of the parameters. It is worth to mention that there is no need to do any plug-in as happened in classical approach.

3.2.2 HPD Predictive Intervals

We use an adaptive generalization of the method used to obtain Highest Posterior Density (HPD) intervals of model parameters, considering predictive distribution

instead of the posterior. Hence the 100 $\gamma\%$ HPD predictive interval for X_{n+1} is defined by $R(\gamma) = (X_L, X_U)$ if

$$P(X_L \leq X_{n+1} \leq X_U) = \sum_{x_{n+1}=X_L}^{X_U} \pi(x_{n+1}|\mathbf{B}_n) \geq K_\gamma,$$

with K_γ being the largest constant such as $P[X_{n+1} \in R(\gamma)] \geq \gamma$.

We can obtain an approximation to the HPD predictive interval for X_{n+1} using the algorithm developed by [1]. After computing the 100 $\gamma\%$ credible intervals

$$\hat{R}_i(\gamma) = (X_{n+1,i}, X_{n+1,i+[m\gamma]}), \quad 1 \leq i \leq m - [m\gamma],$$

where $[m\gamma]$ is the integer part of $m\gamma$, the 100 $\gamma\%$ HPD interval, denoted by $\hat{R}(\gamma)$ is the one with the smallest amplitude.

4 Simulation Study

In this section we study the performance of the above classical and Bayesian procedures with count time series simulated by choosing various combinations of the parameters of INBL(1,0,1,1) model under stationarity conditions.

4.1 Inference

Through the simulation study we want to highlight the following issues: (a) how the results depend on the underlying bilinear parameter β ; (b) what is the impact of sample size on the simulation results, and (c) what is the influence of the variance of the innovation process.

We simulated samples from INBL(1,0,1,1) model of length $n = 50, 100$, and 500 with 100 independent replicates.² In the absence of prior information we consider non-informative priors letting the hyperparameters equal to 0.0001. The MCMC algorithm was used with starting values based on the CLS-estimates and was run with 31,000 iterations in total, the 11,000 initial burn-in iterations were discarded and only the 20th value of the last iterations is kept to reduce the autocorrelation within the chain. Nevertheless, the stationarity of the chain and the convergence of the algorithm were duly analyzed with the usual diagnostic tests, respectively,

²The computation of Bayesian estimates is very demanding in terms of CPU time. Using an Inter Core i5 @ 1.8 GHz-4 GB RAM, the average computation time for producing the estimates of the parameters for samples with size $n = 100$ is approximately 3 days.

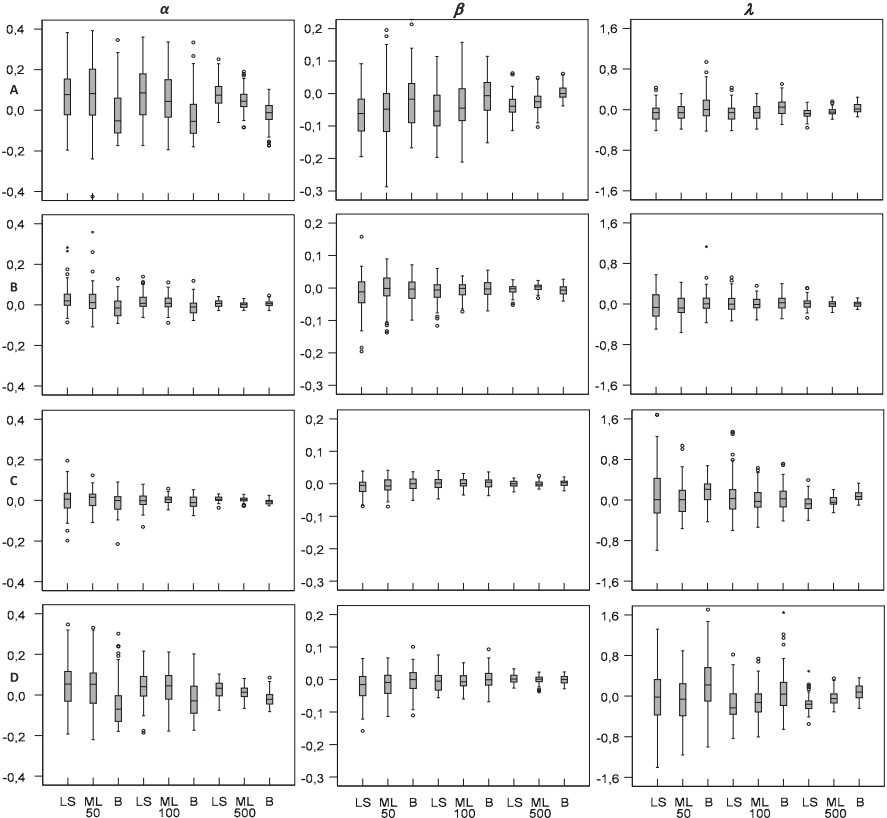


Fig. 2 Boxplots of the biases for $\theta = (\alpha, \beta, \lambda)$ in models A:=(0.2,0.2,1), B:= (0.1,0.6,1), C:=(0.7,0.2,1) and D:=(0.2,0.2,2), with $n = 50, 100, 500$

[8] and [6] tests, which are available in package CODA. Figure 2 displays the boxplots of the biases of CLS-estimates, CML-estimates, and Bayesian estimates for θ , considering each model and the variation of sample size. Concerning the estimation of α a closer look at the figure reveals that classical estimators tend to overestimate the autoregressive parameter, in particular for the models A and D. On the other hand, β is underestimated by any methodology in models A, B, and D. Nevertheless considering all the parameters β is the one that is estimated more accurately. A comparison of the dispersion for the classical and Bayesian estimators shows the similarity for both small and large sample sizes. An important conclusion is that the value of the underlying bilinear parameter does not seem to interfere with the quality of the point estimates for this model. However the variance of the innovation process has large biases, which increases when the theoretical value of λ parameter rises, showing a significant degree of variability. As expected, both the bias and the skewness are also reduced when the sample size increases.

4.2 Prediction

To compare and analyze the different h -step-ahead predictors previously mentioned in Sect. 3 we simulated samples with sizes $n = 50, 100, 200$ from model (1). In order to obtain point or interval forecasts from classical approach the CML estimates were plugged in in (2) or in the predictive probability functions. To obtain Bayesian predictions, we used the expression (4) and Tanner algorithm [13]. It is worth to notice that the forecast performance depends on one hand, on the difference between x_n and x_{n+h} , $h \geq 1$, similarly to what happens in INAR(1) model as described by [11] and on the other hand, on the prediction errors e_n . This situation is illustrated as follows: the forecasts of $x_{168} = 8$ are $\hat{x}_{168, \text{CML}} = 12.684$, $\hat{x}_{168, B} = 13.010$, $\hat{M}_{168}^{(\text{CML})} = \hat{M}_{168}^{(B)} = 13$ are closer to $x_{167} = 13$, when $\alpha = 0.7$, $\beta = 0.2$, $\lambda = 1$.

In Fig. 3 the h -step-ahead predictions, considering two particular sets of parameters and $n = 100$, are plotted. These plots indicate that the obtained results for the predictions using the classical approach with CML-estimates and the Bayesian methodology are very similar. It must be emphasized that these predictions are closed to the limit values given by (3), corresponding to 5.33 for $\theta = (0.1, 0.6, 1)$ and to 12 when $\theta = (0.7, 0.2, 1)$. Figures 4 and 5 represent the amplitude means of the prediction intervals or the HPD predictive intervals for the future value and the frequencies of the simulated X_{n+1} belonging to the prediction interval, respectively. We observe that in general classical prediction intervals based on CML-estimates present smaller amplitude means than the Bayesian correspondents. Another

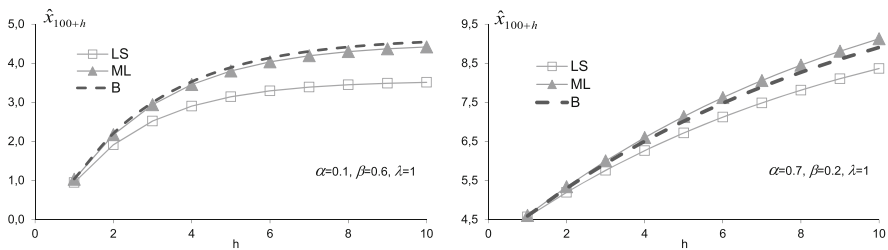


Fig. 3 h -step-ahead predictions for future observations

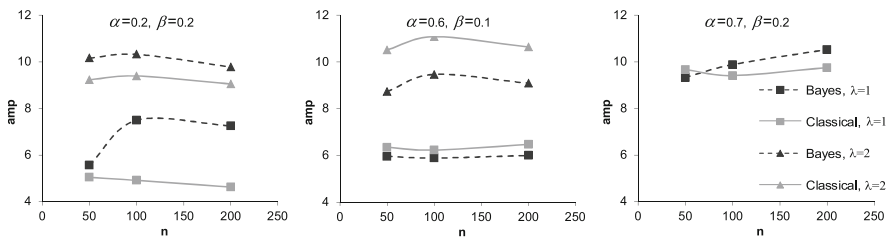


Fig. 4 Means of the prediction interval amplitudes of X_{n+1}

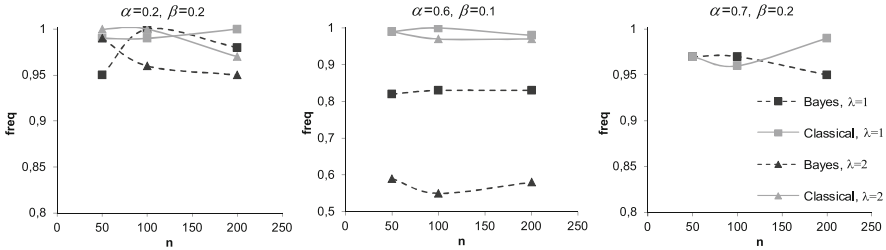


Fig. 5 Frequencies of X_{n+1} belonging to the prediction intervals

important feature exhibited is that the percentage of the simulated observation X_{n+1} belonging to the classical interval is greater than 96.

5 Application to Real Data

In this section, we illustrate the modelling procedure with the motivating examples presented in Fig. 1. It could be pointed out that both data sets are asymmetric with significant overdispersion in E.coli data, with empirical mean and variance being 2.3 and 13.03, respectively.

We should check the adequacy of the distributional assumptions of the model. For this purpose we use the nonrandomized version of PIT histogram, proposed by [2] (see [9], for further models evaluation based in its predictive performance). The graphical tools represented in Fig. 6 are the PIT histograms and the mean PIT charts applied to the data sets. From left to right, the PIT histograms are U-shaped and uniform indicating underdispersed and well-calibrated predictive distributions, respectively. These plots indicate that the probability structure addressed to the INBL(1,0,1,1) is misspecified in the E.coli data despite the Pearson residuals exhibit mean 0.0002, variance 0.9999 and no significant serial correlation. Results of the parameter estimates for the meningitis data are presented in Table 1. However the bilinear component β in the model seems to be very small, which may question its interest in the model. Finally, in order to evaluate and compare the different

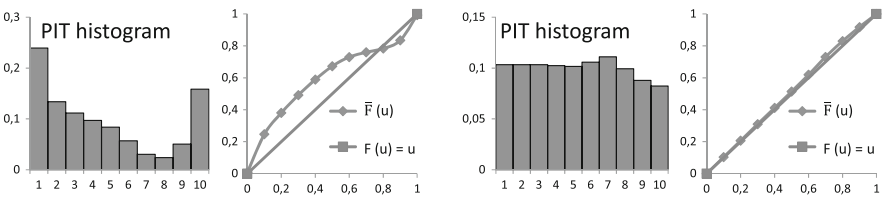


Fig. 6 PIT histograms and charts of mean PIT (denoted by $\bar{F}(u)$), applied to E.coli data (left) and meningitis data (right)

Table 1 Estimated model for the meningitis data

Data	$\hat{\alpha}_{CLS}$	$\hat{\beta}_{CLS}$	$\hat{\lambda}_{CLS}$	$\hat{\alpha}_{CML}$	$\hat{\beta}_{CML}$	$\hat{\lambda}_{CML}$	$\hat{\alpha}_B$	$\hat{\beta}_B$	$\hat{\lambda}_B$
Meningitis data	0.151	0.027	0.296	0.201	0.026	0.288	0.181	0.029	0.290

Table 2 h -step-ahead
predictions using meningitis
data

h	1	2	3
$\hat{X}_{140+h}^{(CML)}$	0.296	0.357	0.369
$\hat{X}_{140+h}^{(B)}$	0.308	0.361	0.368
x_{140+h}	0	0	0

predictions methodologies, h -step-ahead forecasts ($h = 1, 2, 3$) are produced for the last 3 observations. From inspection of Table 2 it can be seen that the forecasts obtained by classical and Bayesian approaches are very similar (and close to the real values) which is not a surprising result since the correspondent parameter estimates are very closed. Regarding the predictions one step ahead of x_{141} and using the coherent predictions given by the medians $\hat{M}_{141}^{(CML)}$ or $\hat{M}_{141}^{(B)}$ we obtain the value 1.

6 Concluding Remarks

In this work classical and Bayesian approaches to time series analysis and forecasting are applied to INBL (1,0,1,1) model. However much of the work for INBL processes remains to be done. We can point out some issues that are still open questions: invertibility conditions and the probabilistic structure of the process. This class of models, due to the cross term, can generate extreme observations and hence is suitable for modelling series of counts showing heavy tailed phenomena. However these features increase the difficulty in obtaining good predictions. Throughout this work we have seen that statistical modelling of INBL processes leads to likelihood functions based on convolutions. The difficulty of computing these functions exactly points towards the development of likelihood estimation by saddlepoint approximation, as suggested by [10], and the improvement of MCMC algorithms.

Acknowledgements This work was supported by Portuguese funds through the CIDMA—Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (“FCT—Fundação para a Ciência e a Tecnologia”), within project UID/MAT/04106/2013.

References

1. Chen, M.-H., Shao, Q.-M.: Monte Carlo estimation of Bayesian credible and HPD intervals. J. Comput. Graph. Stat. **8**, 69–92 (1999)

2. Czado, C., Gneiting, T., Held, L.: Predictive model assessment for count data. Biometrics **65**, 1254–1261 (2009)

3. Doukhan, P., Latour, A., Oraichi, D.: A simple integer-valued bilinear times series models. *Adv. Appl. Probab.* **38**, 559–578 (2006)
4. Drost, F.C., van den Akker, R., Werker, B.J.: Note on integer-valued bilinear time series models. *Stat. Probab. Lett.* **78**, 992–996 (2008)
5. Freeland, R.K., McCabe, B.P.M.: Forecasting discrete valued low count time series. *Int. J. Forecast.* **20**, 427–434 (2004)
6. Geweke, J.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F. (eds.) *Bayesian Statistics*, vol. 4, pp. 169–193. Oxford University Press, Oxford (1992)
7. Gilks, W.R., Best, N.G., Tan, K.K.C.: Adaptive rejection metropolis sampling within Gibbs sampling. *Appl. Stat.* **44**, 455–472 (1995)
8. Heidelberger, P., Welch, P.D.: Simulation run length control in the presence of an initial transient. *Oper. Res.* **31**, 1109–1144 (1983)
9. Jung, R.C., Tremayne, A.R.: Useful models for time series of counts or simply wrong ones? *AStA Adv. Stat. Anal.* **95**, 59–91 (2011)
10. Pedeli, X., Davison, A.C., Fokianos, K.: Likelihood estimation for the INAR(p) model by saddlepoint approximation. *J. Am. Stat. Assoc.* **110**, 1229–1236 (2015)
11. Silva, N., Pereira, I., Silva, M.E.: Forecasting in INAR (1) model. *REVSTAT - Stat. J.* **7**, 119–134 (2009)
12. Steutel, F., van Harn, K.: Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**, 893–899 (1979)
13. Tanner, M.A.: *Tools for Statistical Inference*. Springer, New York (1996)